cisco Live!

#CiscoLive

The bridge to possible

Monitoring and Troubleshooting vPC and VXLAN EVPN in a Datacenter Environment

Cameron Buie, Technical Consulting Engineer

BRKDCN-2106

#CiscoLive

Cisco Webex App

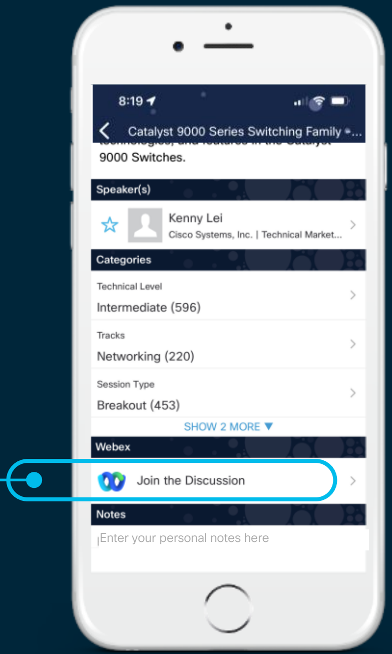
Questions?

Use Cisco Webex App to chat
with the speaker after the session

How

1. Find this session in the Cisco Live Mobile App
2. Click “Join the Discussion”
3. Install the Webex App or go directly to the Webex space
4. Enter messages/questions in the Webex space

Webex spaces will be moderated
by the speaker until June 17, 2022.



8:19

Catalyst 9000 Series Switching Family ...

9000 Switches.

Speaker(s)

Kenny Lei
Cisco Systems, Inc. | Technical Market...

Categories

Technical Level
Intermediate (596)

Tracks
Networking (220)

Session Type
Breakout (453)

SHOW 2 MORE ▼

Webex

Join the Discussion

Notes

Enter your personal notes here

https://ciscolive.ciscoevents.com/ciscolivebot/#BRKDCN-2106

Agenda

- **Introduction**
- Monitoring vPC and VXLAN EVPN
- Troubleshooting Commonly seen problems by TAC
- Conclusion
- QA

Agenda

- Introduction
- **Monitoring vPC and VXLAN EVPN**
- Troubleshooting Commonly seen problems by TAC
- Conclusion
- QA

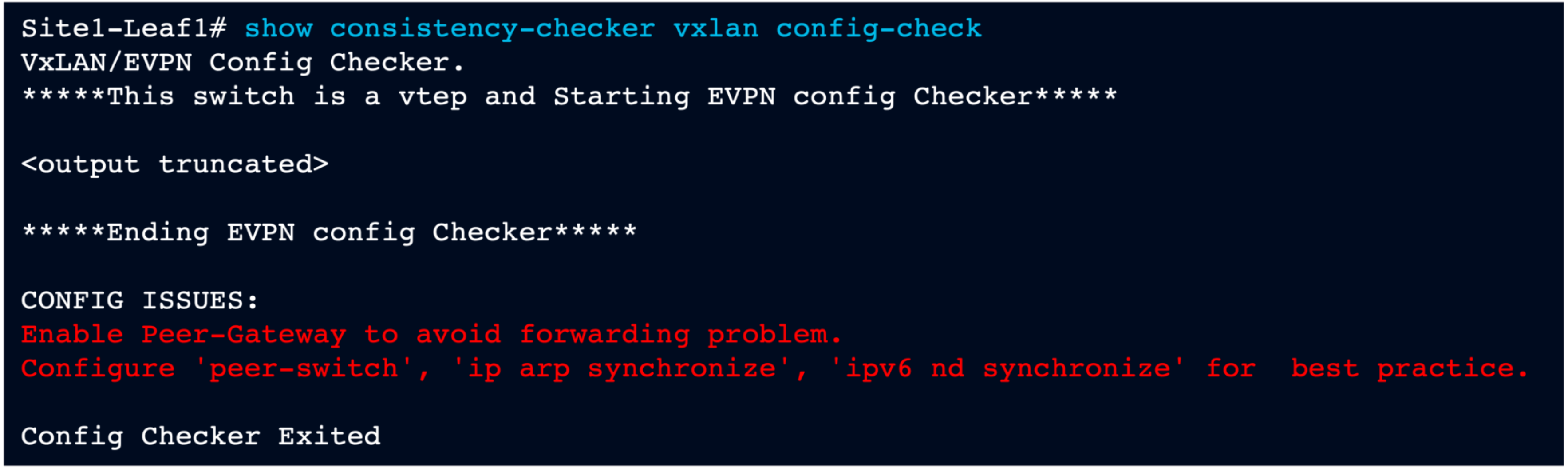
Monitoring vPC and VXLAN EVPN

- **Consistency Checkers**
- ICAM Scale Monitoring

Consistency Checkers

- Configuration Check
  - Looks for best practices
  - Checks for common misconfigurations

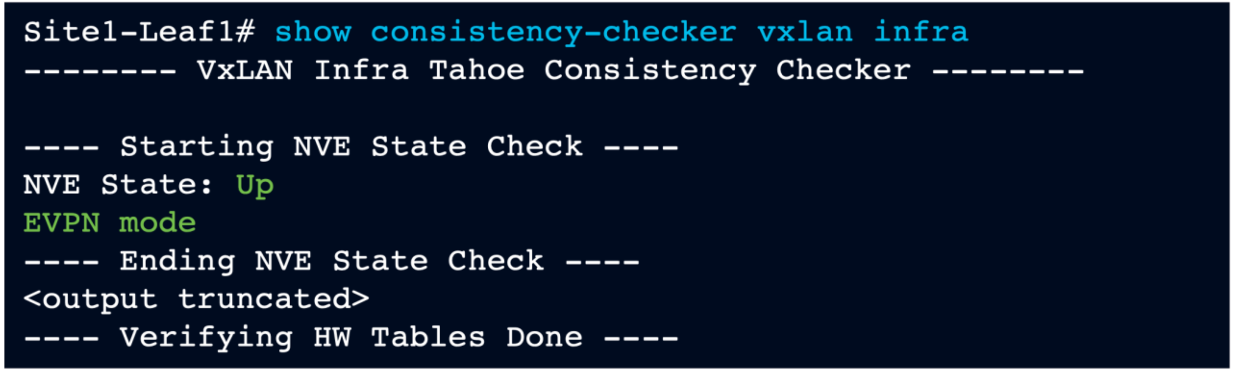
Note: VXLAN Consistency-checkers are not supported for TRM, Flood and Learn and Asymmetric/Downstream



```
Site1-Leaf1# show consistency-checker vxlan config-check
VxLAN/EVPN Config Checker.
*****This switch is a vtep and Starting EVPN config Checker*****

<output truncated>

*****Ending EVPN config Checker*****

CONFIG ISSUES:
Enable Peer-Gateway to avoid forwarding problem.
Configure 'peer-switch', 'ip arp synchronize', 'ipv6 nd synchronize' for  best practice.

Config Checker Exited
```

Consistency Checkers

- Infrastructure Check
  - Control Plane consistency
  - VXLAN state in software and hardware
  - Multicast check for BUM traffic



```
Site1-Leaf1# show consistency-checker vxlan infra
-------- VxLAN Infra Tahoe Consistency Checker --------

---- Starting NVE State Check ----
NVE State: Up
EVPN mode
---- Ending NVE State Check ----
<output truncated>
---- Verifying HW Tables Done ----
```

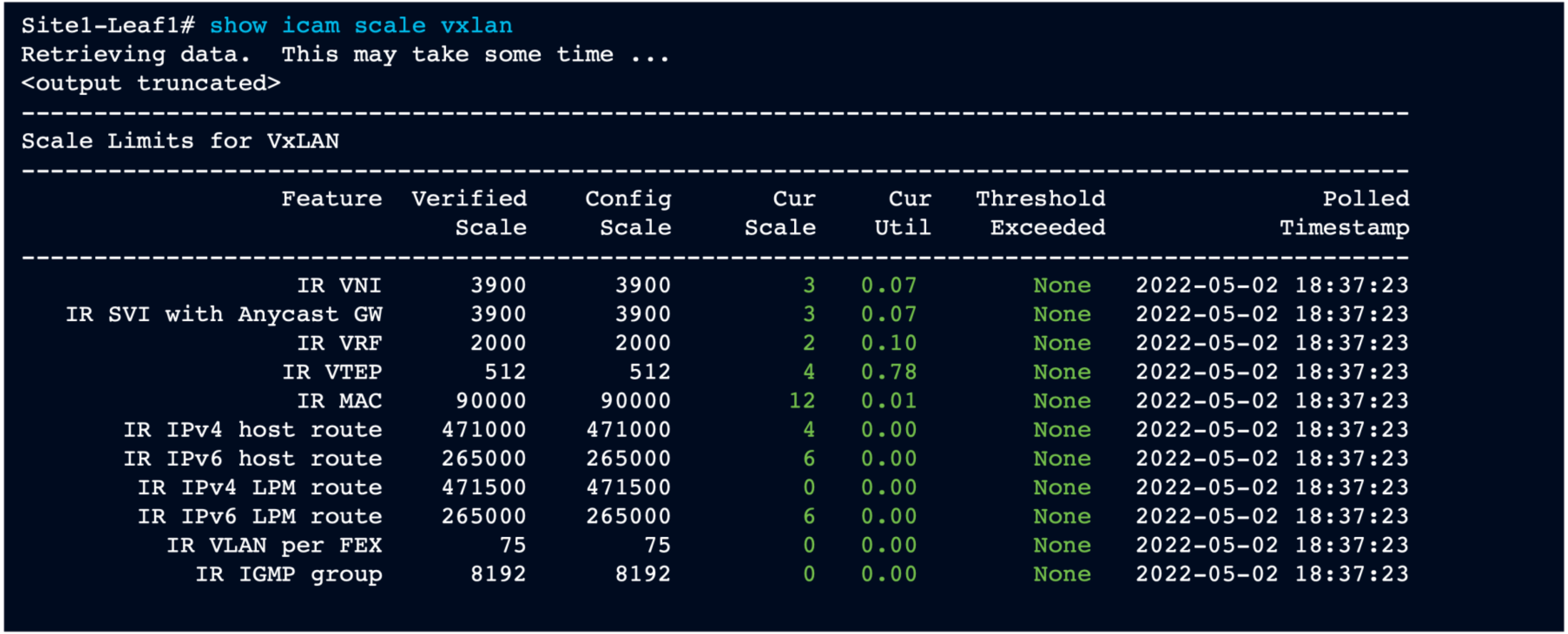
Monitoring vPC and VXLAN EVPN

- Consistency Checkers
- **ICAM Scale Monitoring**

ICAM Scale Checking

- Feature is on by default – 9.3(5) and later
- Syslog at high utilization
- Periodic monitoring configurable
  - Stored history size configurable
  - Entire output saved each interval
- Checks scale of multiple components
  - Not only for VXLAN EVPN
  - Can monitor fib and ACL TCAM, system, memory, and more

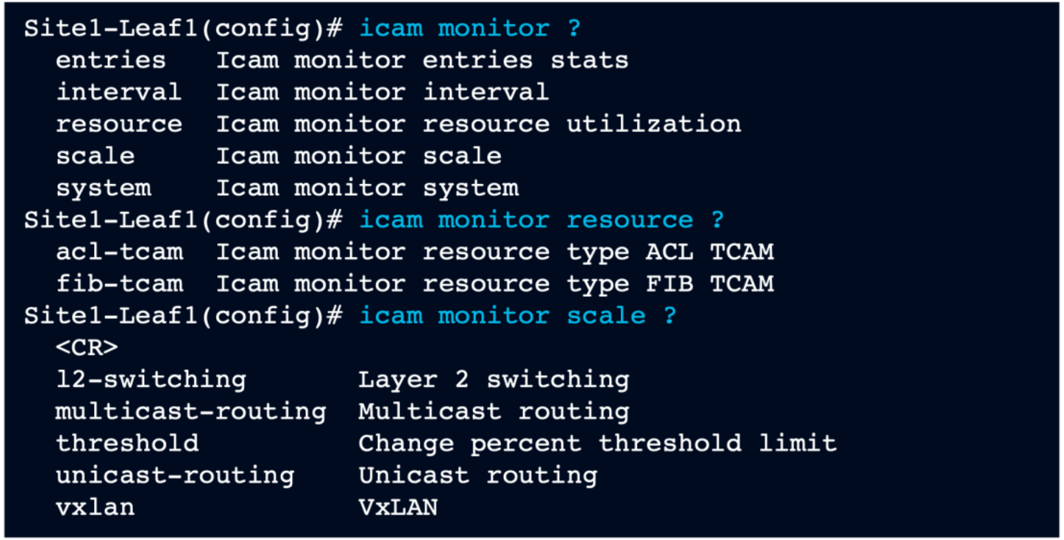
ICAM VXLAN Scale Report



```
Site1-Leaf1# show icam scale vxlan
Retrieving data.  This may take some time ...
<output truncated>
------------------------------------------------------------------------------------------------
Scale Limits for VxLAN
------------------------------------------------------------------------------------------------
                  Feature  Verified    Config       Cur     Cur   Threshold               Polled
                              Scale     Scale     Scale    Util    Exceeded            Timestamp
------------------------------------------------------------------------------------------------
                   IR VNI      3900      3900         3   0.07        None   2022-05-02 18:37:23
   IR SVI with Anycast GW      3900      3900         3   0.07        None   2022-05-02 18:37:23
                   IR VRF      2000      2000         2   0.10        None   2022-05-02 18:37:23
                  IR VTEP       512       512         4   0.78        None   2022-05-02 18:37:23
                   IR MAC     90000     90000        12   0.01        None   2022-05-02 18:37:23
       IR IPv4 host route    471000    471000         4   0.00        None   2022-05-02 18:37:23
       IR IPv6 host route    265000    265000         6   0.00        None   2022-05-02 18:37:23
        IR IPv4 LPM route    471500    471500         0   0.00        None   2022-05-02 18:37:23
        IR IPv6 LPM route    265000    265000         6   0.00        None   2022-05-02 18:37:23
          IR VLAN per FEX        75        75         0   0.00        None   2022-05-02 18:37:23
            IR IGMP group      8192      8192         0   0.00        None   2022-05-02 18:37:23
```

Configuring ICAM Scale Monitoring

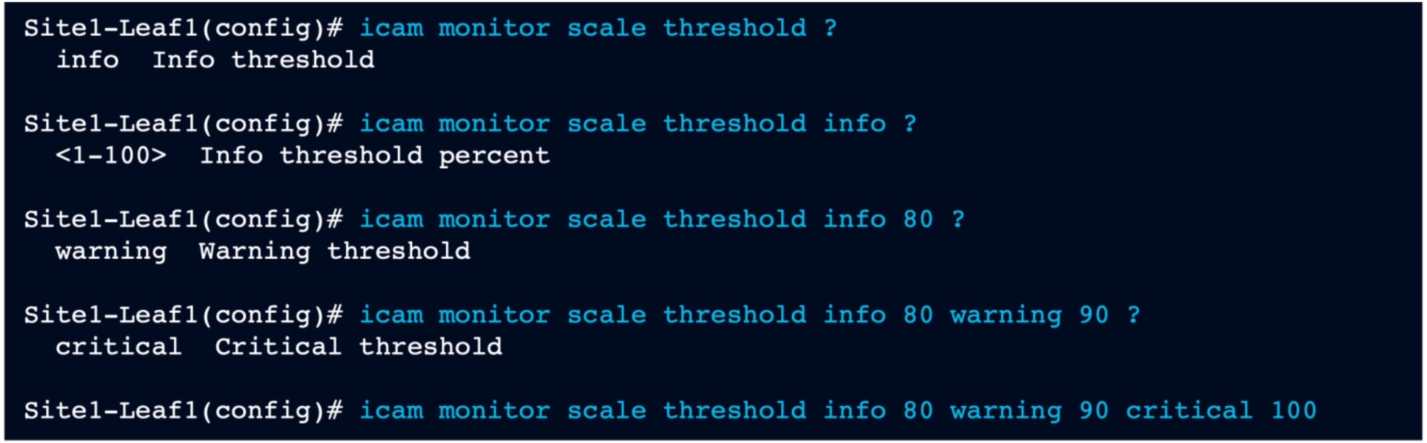
- Multiple options to choose from
  - Can monitor fib and ACL tcam
  - Can monitor multiple components at the same time



```
Site1-Leaf1(config)# icam monitor ?
  entries   Icam monitor entries stats
  interval  Icam monitor interval
  resource  Icam monitor resource utilization
  scale     Icam monitor scale
  system    Icam monitor system
Site1-Leaf1(config)# icam monitor resource ?
  acl-tcam  Icam monitor resource type ACL TCAM
  fib-tcam  Icam monitor resource type FIB TCAM
Site1-Leaf1(config)# icam monitor scale ?
  <CR>
  l2-switching       Layer 2 switching
  multicast-routing  Multicast routing
  threshold          Change percent threshold limit
  unicast-routing    Unicast routing
  vxlan              VxLAN
```

Configuring ICAM alert thresholds

- Alert thresholds are configurable
  - Defaults are 80% informational, 90% warning, 100% critical
- Alert will come via a syslog message.



```
Site1-Leaf1(config)# icam monitor scale threshold ?
  info   Info threshold

Site1-Leaf1(config)# icam monitor scale threshold info ?
  <1-100>   Info threshold percent

Site1-Leaf1(config)# icam monitor scale threshold info 80 ?
  warning   Warning threshold

Site1-Leaf1(config)# icam monitor scale threshold info 80 warning 90 ?
  critical   Critical threshold

Site1-Leaf1(config)# icam monitor scale threshold info 80 warning 90 critical 100
```

Configuring ICAM Scale Monitoring History

- History size is configurable with a default of 168 entries
- Poll interval is configurable with a default of every 2 hours
- Show icam scale vxlan history <entry #> to view historical report
- Historic logs are not enabled until “icam monitor scale” is configured.

```
Site1-Leaf1(config)# icam monitor interval ?
  <1-24>   Icam monitor interval in hours

Site1-Leaf1(config)# icam monitor interval 1 ?
  history   Icam monitor history

Site1-Leaf1(config)# icam monitor interval 1 history ?
  <168-1344>   Number of intervals to keep in icam monitor history
```

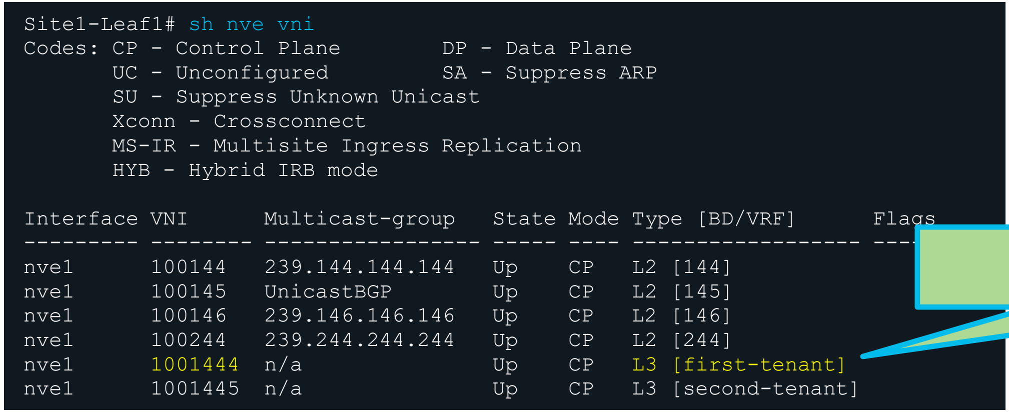
Agenda

- Introduction
- Monitoring vPC and VXLAN EVPN
- **Troubleshooting Commonly seen problems by TAC**
- Conclusion
- QA

Troubleshooting Commonly seen problems by TAC

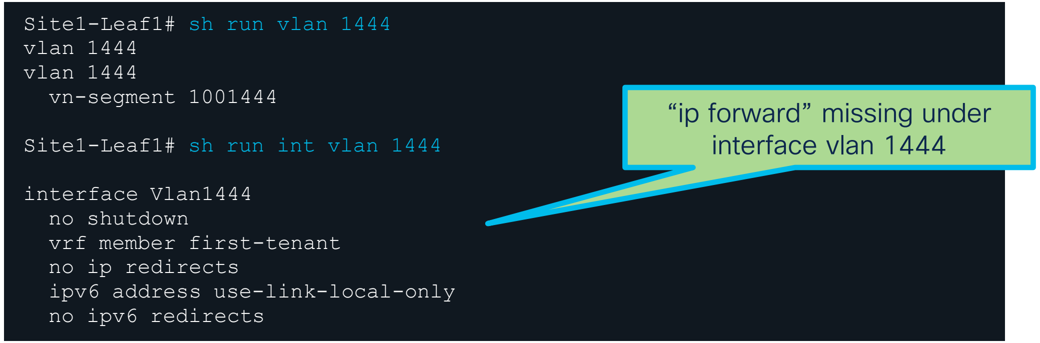
- **ip forward missing**
- Underlay sub-interfaces
- ARP suppression without anycast GW
- eBGP peering to host
- MTU

ip forward missing



```
Site1-Leaf1# sh nve vni
Codes: CP - Control Plane        DP - Data Plane
       UC - Unconfigured         SA - Suppress ARP
       SU - Suppress Unknown Unicast
       Xconn - Crossconnect
       MS-IR - Multisite Ingress Replication
       HYB - Hybrid IRB mode

Interface VNI      Multicast-group   State Mode Type [BD/VRF]      Flags
--------- -------- ----------------- ----- ---- ------------------ ---
nve1      100144   239.144.144.144   Up    CP   L2 [144]
nve1      100145   UnicastBGP        Up    CP   L2 [145]
nve1      100146   239.146.146.146   Up    CP   L2 [146]
nve1      100244   239.244.244.244   Up    CP   L2 [244]
nve1      1001444  n/a               Up    CP   L3 [first-tenant]
nve1      1001445  n/a               Up    CP   L3 [second-tenant]
```

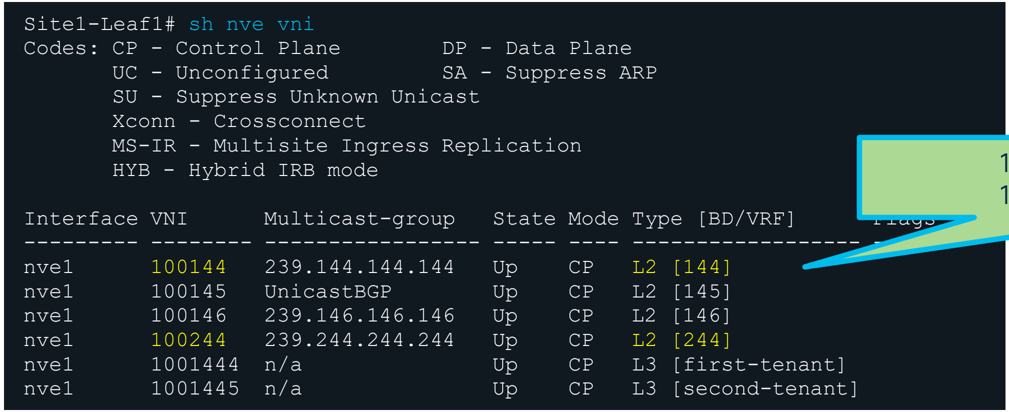
1001444 is the L3 VNI for vrf first-tenant



```
Site1-Leaf1# sh run vlan 1444
vlan 1444
vlan 1444
  vn-segment 1001444

Site1-Leaf1# sh run int vlan 1444

interface Vlan1444
  no shutdown
  vrf member first-tenant
  no ip redirects
  ipv6 address use-link-local-only
  no ipv6 redirects
```

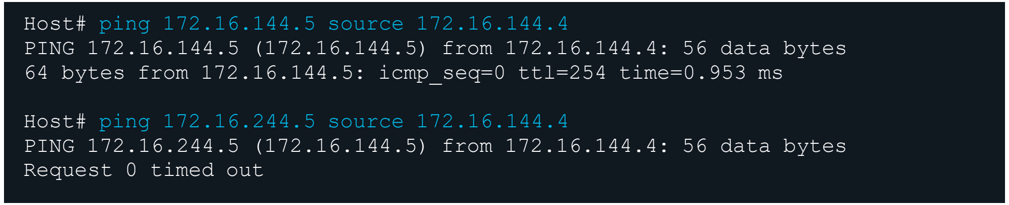
“ip forward” missing under interface vlan 1444

ip forward missing example



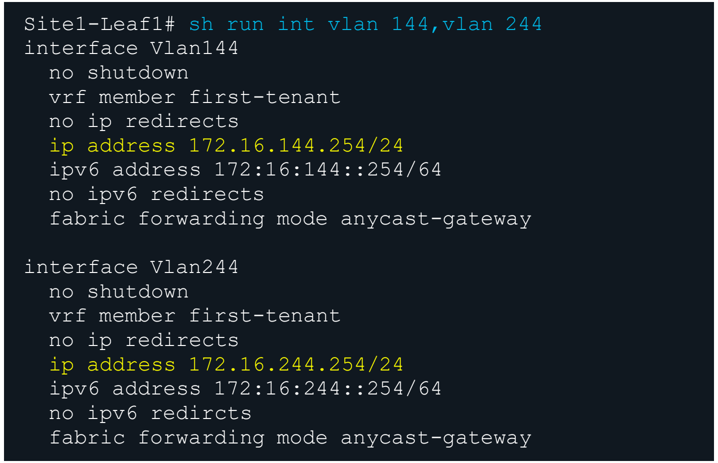
```
Site1-Leaf1# sh nve vni
Codes: CP - Control Plane        DP - Data Plane
       UC - Unconfigured         SA - Suppress ARP
       SU - Suppress Unknown Unicast
       Xconn - Crossconnect
       MS-IR - Multisite Ingress Replication
       HYB - Hybrid IRB mode

Interface VNI      Multicast-group   State Mode Type [BD/VRF]      Flags
--------- -------- ----------------- ----- ---- ------------------ -
nve1      100144   239.144.144.144   Up    CP   L2 [144]
nve1      100145   UnicastBGP        Up    CP   L2 [145]
nve1      100146   239.146.146.146   Up    CP   L2 [146]
nve1      100244   239.244.244.244   Up    CP   L2 [244]
nve1      1001444  n/a               Up    CP   L3 [first-tenant]
nve1      1001445  n/a               Up    CP   L3 [second-tenant]
```

100144 is the L2 VNI for vlan 144
100244 is the L2 VNI for vlan 244



```
Host# ping 172.16.144.5 source 172.16.144.4
PING 172.16.144.5 (172.16.144.5) from 172.16.144.4: 56 data bytes
64 bytes from 172.16.144.5: icmp_seq=0 ttl=254 time=0.953 ms

Host# ping 172.16.244.5 source 172.16.144.4
PING 172.16.244.5 (172.16.144.5) from 172.16.144.4: 56 data bytes
Request 0 timed out
```

Same VNI traffic is working properly. Routed traffic between VNI's is broken



```
Site1-Leaf1# sh run int vlan 144,vlan 244
interface Vlan144
  no shutdown
  vrf member first-tenant
  no ip redirects
  ip address 172.16.144.254/24
  ipv6 address 172:16:144::254/64
  no ipv6 redirects
  fabric forwarding mode anycast-gateway

interface Vlan244
  no shutdown
  vrf member first-tenant
  no ip redirects
  ip address 172.16.244.254/24
  ipv6 address 172:16:244::254/64
  no ipv6 redircts
  fabric forwarding mode anycast-gateway
```

ip forward missing

```
Site1-Leaf1# conf t
Enter configuration commands, one per line. End with CNTL/Z.
Site1-Leaf1(config)# int vlan 1444
Site1-Leaf1(config-if)# ip forward
Site1-Leaf1(config-if)# end
```

Add ip forward on interface vlan 1444

```
Host# ping 172.16.144.5 source 172.16.144.4
PING 172.16.144.5 (172.16.144.5) from 172.16.144.4: 56 data bytes
64 bytes from 172.16.144.5: icmp_seq=0 ttl=254 time=0.953 ms

Host# ping 172.16.244.5 source 172.16.144.4
PING 172.16.244.5 (172.16.144.5) from 172.16.144.4: 56 data bytes
64 bytes from 172.16.244.5: icmp_seq=0 ttl=252 time=1.155 ms
```

Both pings now working as expected

- ip forward is required on VRF SVI’s
- Types of flows affected
  - CPU bound traffic (with L3VNI) – DHCP Offer messages, eBGP messages
  - Routed traffic arriving on the L3VNI
  - Not all scenarios listed

Troubleshooting Commonly seen problems by TAC

- ip forward missing
- **Underlay sub-interfaces**
- ARP suppression without anycast GW
- eBGP peering to host
- MTU

Underlay sub-interfaces

- Sub-interfaces for uplinks are not supported.
- VTEP's do not support VXLAN traffic over sub-interfaces.
- Starting in NX-OS 9.3(5) VXLAN encapsulated traffic can flow over a parent interface if sub-interfaces are configured.

Troubleshooting Commonly seen problems by TAC

- ip forward missing
- Underlay sub-interfaces
- **ARP suppression without anycast GW**
- eBGP peering to host
- MTU

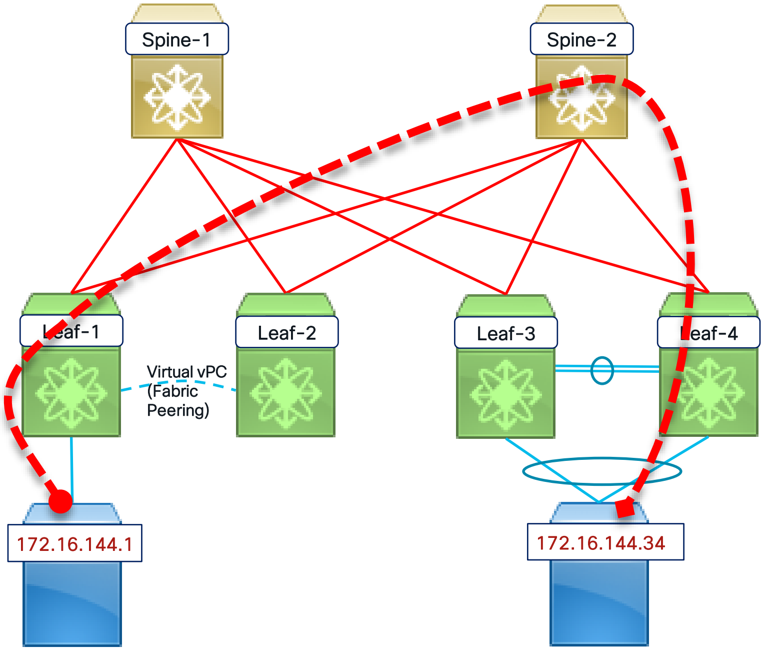
ARP Requests – No SVI on L2VNI, ARP suppression is configured on the L2VNI

```
Host# ping 172.16.144.34 source 172.16.144.1
PING 172.16.144.34 (172.16.144.34) from 172.16.144.1: 56 data bytes
Request 0 timed out
Request 1 timed out
Request 2 timed out
Request 3 timed out
Request 4 timed out
```

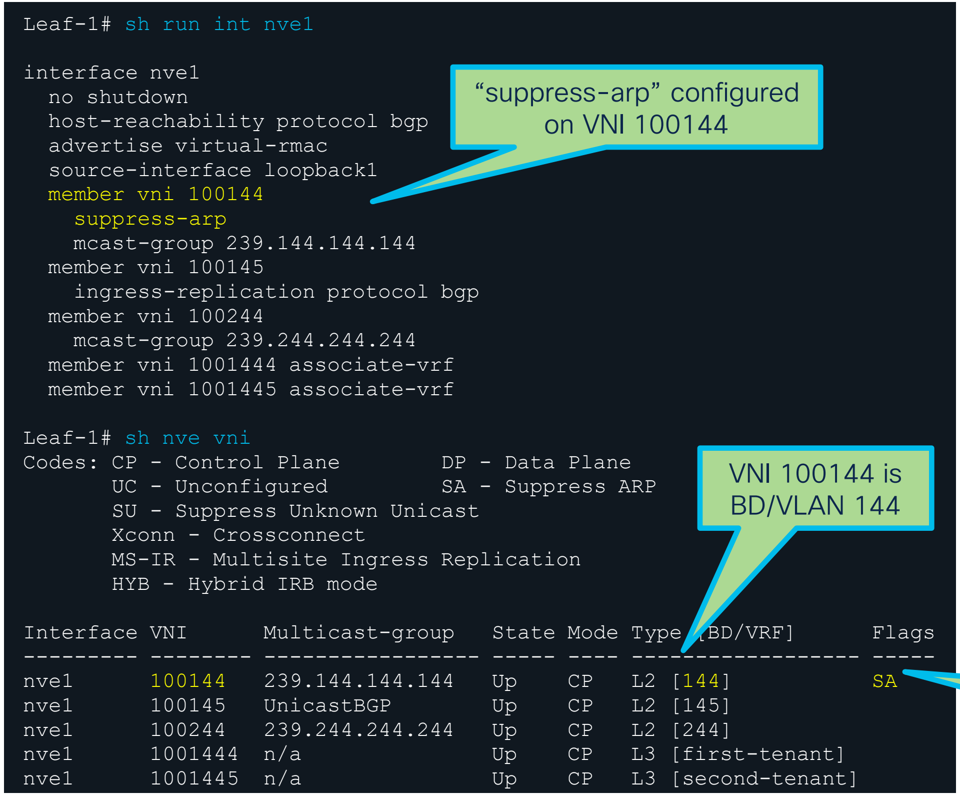
Pings between hosts failing on the same VNI

```
Host# sh ip arp 172.16.144.34

Flags: * - Adjacencies learnt on non-active FHRP router
       + - Adjacencies synced via CFSoE
       # - Adjacencies Throttled for Glean
       CP - Added via L2RIB, Control plane Adjacencies
       PS - Added via L2RIB, Peer Sync
       RO - Re-Originated Peer Sync Entry
       D - Static Adjacencies attached to down interface

IP ARP Table for context default
Total number of entries: 1
Address         Age       MAC Address     Interface       Flags
172.16.144.34   00:00:15  INCOMPLETE      Vlan144
```

ARP incomplete on the host



ARP Requests – No SVI on L2VNI, ARP suppression is configured on the L2VNI

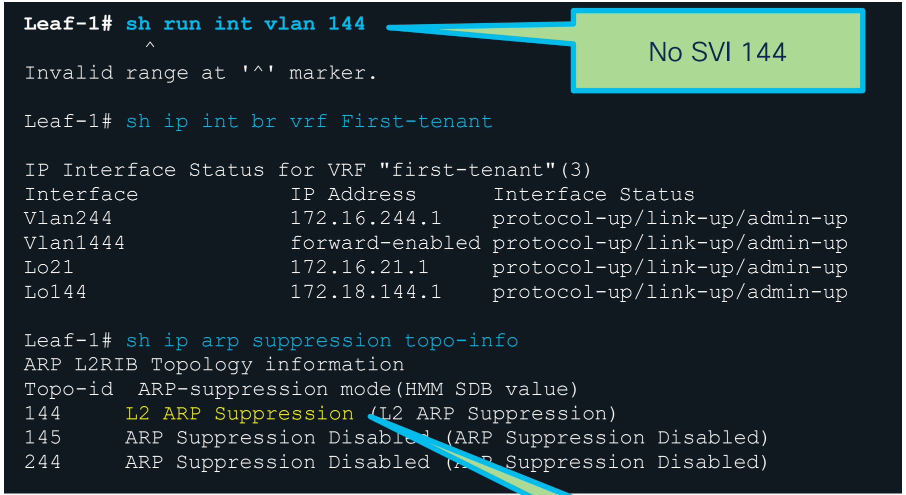


```
Leaf-1# sh run int nve1

interface nve1
  no shutdown
  host-reachability protocol bgp
  advertise virtual-rmac
  source-interface loopback1
  member vni 100144
    suppress-arp
    mcast-group 239.144.144.144
  member vni 100145
    ingress-replication protocol bgp
  member vni 100244
    mcast-group 239.244.244.244
  member vni 1001444 associate-vrf
  member vni 1001445 associate-vrf

Leaf-1# sh nve vni
Codes: CP - Control Plane        DP - Data Plane
       UC - Unconfigured         SA - Suppress ARP
       SU - Suppress Unknown Unicast
       Xconn - Crossconnect
       MS-IR - Multisite Ingress Replication
       HYB - Hybrid IRB mode

Interface VNI      Multicast-group   State Mode Type [BD/VRF]      Flags
--------- -------- ----------------- ----- ---- ------------------ -----
nve1      100144   239.144.144.144   Up    CP   L2 [144]           SA
nve1      100145   UnicastBGP        Up    CP   L2 [145]
nve1      100244   239.244.244.244   Up    CP   L2 [244]
nve1      1001444  n/a               Up    CP   L3 [first-tenant]
nve1      1001445  n/a               Up    CP   L3 [second-tenant]
```

"suppress-arp" configured on VNI 100144

VNI 100144 is BD/VLAN 144

SA = Suppress ARP



```
Leaf-1# sh run int vlan 144
          ^
Invalid range at '^' marker.

Leaf-1# sh ip int br vrf First-tenant

IP Interface Status for VRF "first-tenant"(3)
Interface            IP Address      Interface Status
Vlan244              172.16.244.1    protocol-up/link-up/admin-up
Vlan1444             forward-enabled protocol-up/link-up/admin-up
Lo21                 172.16.21.1     protocol-up/link-up/admin-up
Lo144                172.18.144.1    protocol-up/link-up/admin-up

Leaf-1# sh ip arp suppression topo-info
ARP L2RIB Topology information
Topo-id  ARP-suppression mode(HMM SDB value)
144     L2 ARP Suppression (L2 ARP Suppression)
145     ARP Suppression Disabled (ARP Suppression Disabled)
244     ARP Suppression Disabled (ARP Suppression Disabled)
```

No SVI 144

Suppress ARP active on VLAN 144

cisco Live!

ARP Requests from a Host – No SVI on L2VNI and ARP suppression

- ARP suppression is only supported for a VNI if the VTEP hosts the first hop anycast gateway for this VNI.
- Both the VTEP and the SVI need to be properly configured for the Anycast Gateway operation. This includes:
  - Global Anycast Gateway MAC address configured.
  - Anycast Gateway feature with the virtual IP address on the SVI.

ARP Requests – No SVI on L2VNI, ARP suppression is configured on the L2VNI

```
Leaf4# conf t
Enter configuration commands, one per line. End with CNTL/Z.
Site1-Leaf1(config)# int nve1
Site1-Leaf1(config-if-nve)# member vni 100144
Site1-Leaf1(config-if-nve-vni)# no suppress-arp
Site1-Leaf1(config-if-nve-vni)# shut
Site1-Leaf1(config-if-nve-vni)# no shut
```

```
Leaf4# conf t
Enter configuration commands, one per line. End with CNTL/Z.
Site1-Leaf1(config)# int nve1
Site1-Leaf1(config-if-nve)# member vni 100144
Site1-Leaf1(config-if-nve-vni)# no suppress-arp
Site1-Leaf1(config-if-nve-vni)# shut
Site1-Leaf1(config-if-nve-vni)# no shut
```

shut/no shut the NVE interface after removing suppress arp
Note: This will affect other VLANS/VNI's. This procedure is considered disruptive.

```
Host# ping 172.16.144.34 source 172.16.144.1
PING 172.16.144.34 (172.16.144.34) from 172.16.144.1: 56 data bytes
64 bytes from 172.16.144.34: icmp_seq=0 ttl=254 time=0.953 ms
64 bytes from 172.16.144.34: icmp_seq=1 ttl=254 time=0.505 ms
64 bytes from 172.16.144.34: icmp_seq=2 ttl=254 time=0.445 ms
64 bytes from 172.16.144.34: icmp_seq=3 ttl=254 time=0.442 ms
64 bytes from 172.16.144.34: icmp_seq=4 ttl=254 time=0.587 ms
```

Ping now successful

Troubleshooting Commonly seen problems by TAC

- ip forward missing
- Underlay sub-interfaces
- ARP suppression without anycast GW
- **eBGP peering to host**
- MTU

Leaf eBGP peering

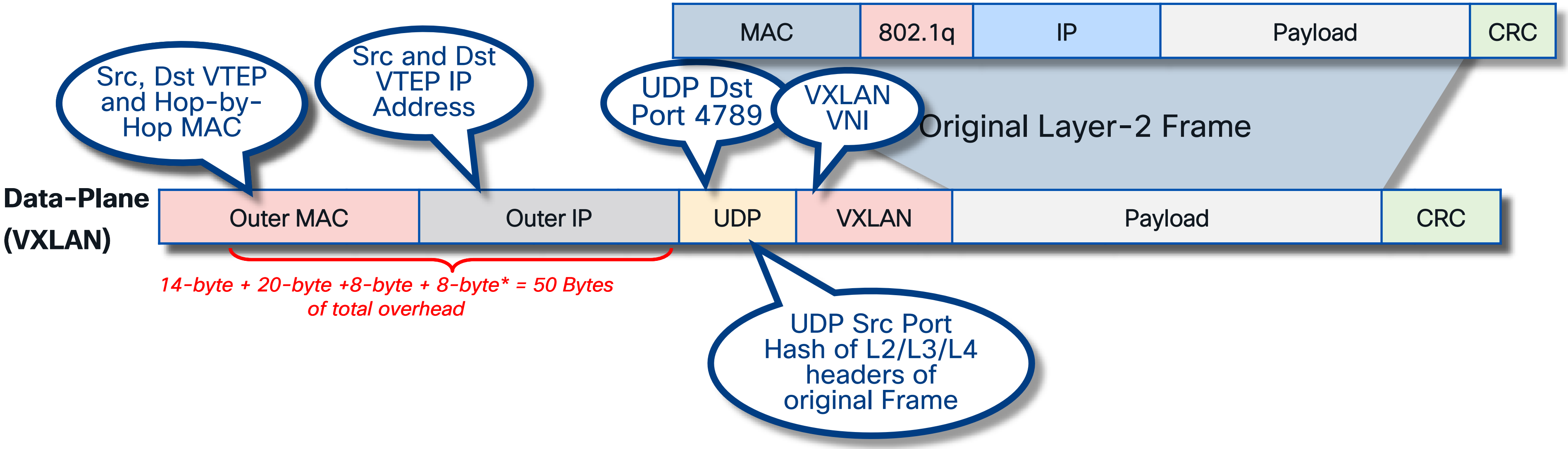
- Use of anycast gateway enabled SVI may result in failure to establish neighborship, due to shared IP among all the leaf switches.
- Dedicated loopback in the tenant vrf must be used to source BGP packets in order to ensure that another leaf switch does not consume the packet due to shared IP among leaf switches.

Troubleshooting Commonly seen problems by TAC

- ip forward missing
- Underlay sub-interfaces
- ARP suppression without anycast GW
- eBGP peering to host
- **MTU**

MTU – VXLAN adds add 50-byte header

- Due to VXLAN encapsulation, the MTU requirement is larger and we need to avoid potential fragmentation.



14-byte + 20-byte +8-byte + 8-byte = 50 Bytes of total overhead*

MTU Issues

- Due to the 50 byte overhead jumbo MTU is required through the spine layer
- Recommended MTU of 9216 on all uplinks and throughout the transport network if possible
- VXLAN traffic does not support fragmentation

Agenda

- Introduction
- Monitoring vPC and VXLAN EVPN
- Troubleshooting Commonly seen problems by TAC
- **Conclusion**
- QA

Agenda

- Introduction
- Monitoring vPC and VXLAN EVPN
- Troubleshooting Commonly seen problems by TAC
- Conclusion
- QA

Technical Session Surveys

- Attendees who fill out a minimum of four session surveys and the overall event survey will get Cisco Live branded socks!

- Attendees will also earn 100 points in the Cisco Live Game for every survey completed.

- These points help you get on the leaderboard and increase your chances of winning daily and grand prizes.

Cisco Learning and Certifications

From technology training and team development to Cisco certifications and learning plans, let us help you empower your business and career. www.cisco.com/go/certs

Pay for Learning with Cisco Learning Credits

(CLCs) are prepaid training vouchers redeemed directly with Cisco.

Learn

Cisco U.
IT learning hub that guides teams and learners toward their goals

Cisco Digital Learning
Subscription-based product, technology, and certification training

Cisco Modeling Labs
Network simulation platform for design, testing, and troubleshooting

Cisco Learning Network
Resource community portal for certifications and learning

Train

Cisco Training Bootcamps
Intensive team & individual automation and technology training programs

Cisco Learning Partner Program
Authorized training partners supporting Cisco technology and career certifications

Cisco Instructor-led and Virtual Instructor-led training
Accelerated curriculum of product, technology, and certification courses

Certify

Cisco Certifications and Specialist Certifications
Award-winning certification program empowers students and IT Professionals to advance their technical careers

Cisco Guided Study Groups
180-day certification prep program with learning and support

Cisco Continuing Education Program
Recertification training options for Cisco certified individuals

Here at the event? Visit us at **The Learning and Certifications lounge at the World of Solutions**

Continue your education

- Visit the Cisco Showcase for related demos
- Book your one-on-one Meet the Engineer meeting
- Attend the interactive education with DevNet, Capture the Flag, and Walk-in Labs
- Visit the On-Demand Library for more sessions at www.CiscoLive.com/on-demand

The bridge to possible

Thank you

cisco Live!

#CiscoLive